

Seamless and multi-resolution energy forecasting

Chenxi Wang^a, Pierre Pinson^{b,c,d}, Yi Wang^{a,*}

^a*Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong SAR, China*

^b*Dyson School of Design Engineering, Imperial College London, London, The United Kingdom*

^c*Department of Technology, Management and Economics, Technical University of Denmark, Denmark*

^d*Halfspace, Denmark*

Abstract

Forecasting is pivotal in energy systems, by providing fundamentals for operation at different horizons and resolutions. Though energy forecasting has been widely studied for capturing temporal information, very few works concentrate on the frequency information provided by forecasts. They are consequently often limited to single-resolution applications (e.g., hourly). Here, we propose a unified energy forecasting framework based on Laplace transform in the multi-resolution context. The forecasts can be seamlessly produced at different desired resolutions without re-training or post-processing. Case studies on both energy demand and supply data show that the forecasts from our proposed method can provide accurate information in both time and frequency domains. Across the resolutions, the forecasts also demonstrate high consistency. More importantly, we explore the operational effects of our produced forecasts in the day-ahead and intra-day energy scheduling. The relationship between (i) errors in both time and frequency domains and (ii) operational value of the forecasts is analysed. Significant operational benefits are obtained.

Keywords: Energy Forecasting, Machine Learning, Laplace Transform, Multi-resolution System Operation

1. Introduction

Energy forecasting is crucial in all segments of the energy industry. The primary goal of energy forecasting is to provide accurate information about both demand and supply in the future, such as electricity load [1] and wind power generation [2]. These forecasts support decision-makers to act optimally [3]. In general, energy forecasting is a time-series prediction problem in which predictions of diverse temporal horizons and resolutions are required for different applications. Here, forecast horizon refers to how far into the future the forecasts are for, while resolution is defined as the time interval between two successive forecast points. For instance, day-ahead energy demand and renewable energy forecasts with hourly or 15-minute resolutions are used by system operators to schedule energy generation in advance [4]. Hours-ahead forecasts at minute resolution are needed for (near) real-time energy balancing [5]. And, forecasts at second resolution levels (nowcasts) can be used for immediate control actions and grid stability analysis [6]. Therefore, energy forecasting at

*Corresponding author

Email addresses: cxwang@eee.hku.hk (Chenxi Wang), p.pinson@imperial.ac.uk (Pierre Pinson), yiwang@eee.hku.hk (Yi Wang)

different horizons and resolutions is fundamental to all aspects of the operation and a management of the entire energy system. At one single resolution level, the energy forecasting problem has been widely studied in the context of both energy demand [7] and supply [8]. Various approaches for either single-step or multi-step forecasting situations can also be easily found in existing literature reviews [9] and books [10]. Despite the prosperity of single-resolution energy forecasting, most of the approaches merely focus on developing a mapping function between relevant temporal features and forecasting targets. However, such methods may be unsuitable for multi-resolution energy forecasting. This is since the frequency information, indicating how intense fluctuations are at various frequencies, may vary depending upon the desired forecasting resolution. Generally, the higher the resolution is, the more detailed frequency information within the energy data should be exhibited [11]. The learned forecasting model for one single resolution can overfit/underfit at other resolutions. Therefore, it may fail to capture the precise frequency information at multiple resolutions.

Even though multiple forecasting models are trained individually for different resolutions [12, 13], there still exists challenges in terms of consistency within the temporal hierarchy of the forecasts [14]. For instance, hourly wind power measurements can be readily deduced by averaging minute-resolution observations over 60-minute periods. From both theoretical and practical points of view, if we forecast at both these resolutions, the forecasts generated should also respect this temporal hierarchy. Multiple models, however, by being naturally unaware of this temporal hierarchy, produce forecasts that are likely to be inconsistent. Therefore, for multi-resolution forecasting, such approaches are no longer suitable. In addition, by changing forecast resolution depending upon user needs, the forecasts need to seamlessly adapt, while also respecting underlying temporal hierarchies.

Some previous works aimed at tackling the above issue. For instance they focus on the post-hoc coordination of multi-resolution forecasts [15, 16], the improved training process of multiple models [17, 18], and continuous-time neural networks for time series [19, 20, 21]. These approaches either try to coordinate the multiple models after/during the training, or purely focus on the dynamics of the time series. They still neglect to reveal the relationship between forecasting resolutions and frequency information. Therefore, from the perspective of methodology, a unified energy forecasting model is needed to seamlessly output the hierarchical forecasts at different resolutions with the corresponding frequency information.

More importantly, energy forecasts with precise frequency information are supposed to have prominent effects on the following decision-making process [22], especially in situations of high resolution. The connection between temporal forecasting errors and resulting operation results is gradually receiving attention [23, 24, 25], but seldom do the existing works explore the effects of errors in the frequency domain. Consequently, from the perspective of empirical analysis, the final operational effects of the forecasts in terms of both time and frequency deserve to be unveiled.

To this end, we propose an innovative and unified energy forecasting framework, inspired by the recently proposed Neural Laplace framework [26], but with the aim to perform seamless and multi-resolution energy forecasting. We refer to it as Hierarchical Neural Laplace (HNL). Given the desired resolutions, the corresponding forecasts can be seamlessly generated without re-training or post-processing. At the desired resolutions, these forecasts can provide high-quality information in both time and frequency domains accordingly. Across the resolutions, the consistent forecasts comply with temporal hierarchies. We also explore the operational effects of our forecasts in representative decision-making processes for energy scheduling. Our analysis of the relationship between temporal errors, frequency errors, and operational costs reveals significant operational benefits.

2. Problem statement

Before introducing our proposed framework, we mathematically state the setting of multi-resolution forecasting in this section.

Given the observed energy data $\mathbf{x} = [x_{t_1}, \dots, x_{t_n}] \in \mathbb{R}^n$ at the sampling resolution f_r , along with the external features $\mathbf{e} \in \mathbb{R}^l$ (e.g. numerical weather predictions, NWP), we aim to produce forecasts in a forecasting horizon at m ascending resolutions f_r^1, \dots, f_r^m ($f_r^m \leq f_r$).

The energy sequence can also be viewed as the discretely sampled points at the resolution f_r from a temporal function $f(t)$, i.e. $x_t = f(t) \in \mathbb{R}$. Therefore, for a given forecasting horizon, the forecasting targets at the resolution f_r^i can be denoted as $\hat{\mathbf{x}}_i = [f'_i(t_{n+1}), \dots, f'_i(t_{n+K_i})] \in \mathbb{R}^{K_i}$, where f'_i is a modeled temporal function for the i th resolution and K_i represents the needed forecast steps for resolution f_r^i . For example, if we would like to forecast the next hour, and the desired resolutions are $f_r^1 = 1/30\text{min}$ and $f_r^2 = 1/15\text{min}$, then $K_1 = f_r^1 \cdot 60\text{min} = 2$ and $K_2 = f_r^2 \cdot 60\text{min} = 4$.

The goal of our proposed framework is to produce multi-resolution forecasts $\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_m$ with one unified model \mathbf{F}_ω :

$$\hat{\mathbf{x}}_i = \mathbf{F}_\omega(\mathbf{x}, \mathbf{e}; f_r^i), \quad \forall i = 1, \dots, m \quad (1)$$

where ω is the set of trainable parameters in the whole model.

3. Hierarchical Neural Laplace

3.1. Overall framework

With the clear aim of multi-resolution forecasting, we briefly introduce the proposed flexible forecasting framework, HNL, in this subsection. Unlike the traditional methods focusing on the time domain, HNL tries to learning the frequency information in the Laplace domain. Then, it utilizes the Inverse Laplace Transform (ILT) to reconstruct the temporal components with different frequency information. According to the desired forecasting resolutions, final energy forecasts can be formed seamlessly from the combination of the temporal components. As shown in Fig. 1, this framework has three core segments: an encoder, multiple Laplace decoders, and an assembler.

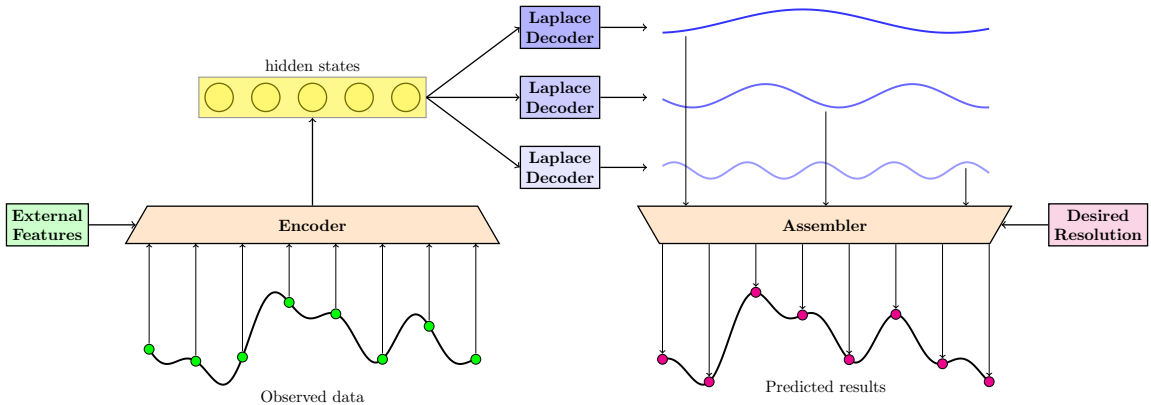


Figure 1: The overview of the proposed Hierarchical Neural Laplace framework

Firstly, the encoder is responsible for transforming observed energy data at some resolution and external features (e.g. numerical weather predictions, NWP) into the high-dimensional representation, i.e., the hidden states. The encoder is model-agnostic, and arbitrary neural networks can be used as encoder, for example, Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), etc.

Then, the encoded hidden states will go through multiple Laplace decoders and produce different temporal components. In this process, there are two sub stages: 1) generating the discrete Laplace function values on pre-determined consecutive frequency bands, and 2) converting these Laplace function values into the temporal components by Fourier-based discrete ILT. The outputs of these decoders are temporal components containing different frequency information, as for the blue curves in Fig. 1.

Finally, given the desired forecasting resolutions, the assembler will fetch groups of temporal components to form the forecasts with precise frequency information. The Shannon sampling theorem, a bridge between data resolution and frequency information, will guide the assembler to decide which groups of components need to be fetched.

In this way, HNL only needs to be trained once on the given data, and flexibly generates the energy forecasts at different resolutions that we are interested in. Details of the framework will be introduced in the following three subsections.

3.2. Encoder

As described above, the encoder ϕ_{ω_e} extracts the observed energy sequence data \mathbf{x} and the external features \mathbf{e} into a hidden state \mathbf{h} :

$$\mathbf{h} = \phi_{\omega_e}(\mathbf{x}, \mathbf{e}), \quad (2)$$

where ω_e represents the trainable parameters in the encoder. Since the encoder is model-agnostic, we do not specify any neural network family here.

3.3. Multiple Laplace Decoders

3.3.1. Laplace transform and ILT

Laplace decoders are the core of our framework, relying on the Laplace transform and ILT. Given a temporal function $f(t)$ of energy data, the corresponding Laplace function $\bar{f}(s)$ through Laplace transform is defined as:

$$\mathcal{L}\{f(t)\} = \bar{f}(s) = \int_0^{\infty} e^{-st} f(t) dt. \quad (3)$$

Similarly, given a Laplace function $\bar{f}(s)$, the temporal function can be inversely transformed as its definition shows:

$$\mathcal{L}^{-1}\{\bar{f}(s)\} = f(t) = \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} e^{st} \bar{f}(s) ds. \quad (4)$$

According to [27], Eq. (4) can be manipulated and discretized through the trapezoid rule:

$$f(t) = \frac{e^{\gamma t}}{\pi} \int_0^{\infty} \Re(\bar{f}(s) e^{i\omega t}) d\omega \quad (5)$$

$$\approx \frac{e^{\gamma t}}{T} \left[\frac{\bar{f}(\gamma)}{2} + \sum_{k=1}^N \Re \left\{ \bar{f} \left(\gamma + \frac{ik\pi}{T} \right) \exp \left(\frac{ik\pi t}{T} \right) \right\} \right], \quad (6)$$

where T and γ are usually set as constants[28], and N controls the number of discrete integral terms, dominating the discretization of ILT.

Further, we rewrite the Eq. (6) with the Euler formula, and obtain the following form:

$$f(t) = \frac{e^{\gamma t}}{T} \left[\frac{\bar{f}(s_0)}{2} + \sum_{k=1}^N \Re \left\{ \bar{f}(s_k) \left(\cos\left(\frac{k\pi t}{T}\right) + i \sin\left(\frac{k\pi t}{T}\right) \right) \right\} \right] \quad (7)$$

where for simplicity, we denote $s_k = \gamma + \frac{ik\pi}{T} \in \mathbb{C}$ and $\bar{\mathbf{f}}(\mathbf{s}) = [\bar{f}(s_0), \bar{f}(s_1), \dots, \bar{f}(s_N)] \in \mathbb{C}^{N+1}$.

We observe that the inversely transformed $f(t)$ explicitly consists of cosine waves with exponential effects at different frequencies:

$$f_c^k = k\pi/2\pi T = k/2T. \quad (8)$$

Consequently, if we are given the discrete Laplace function values $\bar{\mathbf{f}}(\mathbf{s})$, the original $f(t)$ can be reconstructed and an energy forecast $f(t_{n+1})$, for example, can be generated accordingly with explicit frequency information.

3.3.2. Two-stage decoding

Therefore, the general idea of Laplace decoders is first outputting the Laplace function values $\bar{\mathbf{f}}'(s)$ of the energy sequence on the forecasting horizon, and then reconstructing a temporal function $f'(t)$ for forecasting. Here, we use $'$ to distinguish from the real $\bar{\mathbf{f}}(s)$ and $f(t)$ on the forecasting horizon.

In the first stage, we construct p neural networks $g_{\omega_1}, \dots, g_{\omega_p}$ to separately learn the Laplace functions on p consecutive frequency bands:

$$\begin{cases} \bar{\mathbf{f}}'(\mathbf{s}_1) = g_{\omega_1}(\mathbf{s}_1, \mathbf{h}), & \mathbf{s}_1 = [s_0, s_1, \dots, s_{N_1}] \\ \bar{\mathbf{f}}'(\mathbf{s}_2) = g_{\omega_2}(\mathbf{s}_2, \mathbf{h}), & \mathbf{s}_2 = [s_{N_1+1}, s_{N_1+2}, \dots, s_{N_2}] \\ \vdots \\ \bar{\mathbf{f}}'(\mathbf{s}_p) = g_{\omega_p}(\mathbf{s}_p, \mathbf{h}), & \mathbf{s}_p = [s_{N_{p-1}+1}, s_{N_{p-1}+2}, \dots, s_N] \end{cases} \quad (9)$$

where $\omega_1, \dots, \omega_p$ are trainable parameters in these neural networks; $N_1 \leq \dots \leq N_{p-1} (\leq N)$ are $p-1$ split points set manually. All the neural networks share the extracted information of hidden states \mathbf{h} and differ in the corresponding frequency bands. Since $\mathbf{s}_i, \bar{\mathbf{f}}'(\mathbf{s}_i) \in \mathbb{C}^{N_i - N_{i-1} + 1}$ are complex vectors, we split the values of the real and the imaginary parts during inputting and outputting of these neural networks. In other words, the input and output space of g_{ω_i} are respectively $\mathbb{R}^{2(N_i - N_{i-1} + 1) + l}$ and $\mathbb{R}^{2(N_i - N_{i-1} + 1)}$.

In the second stage, the learned Laplace function values $\bar{\mathbf{f}}'(\mathbf{s}_1), \dots, \bar{\mathbf{f}}'(\mathbf{s}_p)$ are fed into discrete ILT as introduced before to reconstruct the temporal function. Further, the reconstructed $f'(t)$ can be presented by the combination of multiple temporal components:

$$f'(t) = \frac{e^{\gamma t}}{T} \left[\underbrace{\frac{\bar{f}(s_0)}{2} + \sum_{k=1}^{N_1} S_k(t)}_{TC_1} + \underbrace{\sum_{k=N_1+1}^{N_2} S_k(t)}_{TC_2} + \dots + \underbrace{\sum_{k=N_{p-1}+1}^{N_p} S_k(t)}_{TC_p} \right], \quad (10)$$

where we denote $S_k(t) = \Re \left\{ \bar{f}'(s_k) \left(\exp\left(\frac{ik\pi t}{T}\right) \right) \right\}$ for simplicity and TC_i represents the i th temporal component. The whole tow-stage decoding of these multiple Laplace decoders is demonstrated by Fig. 2.

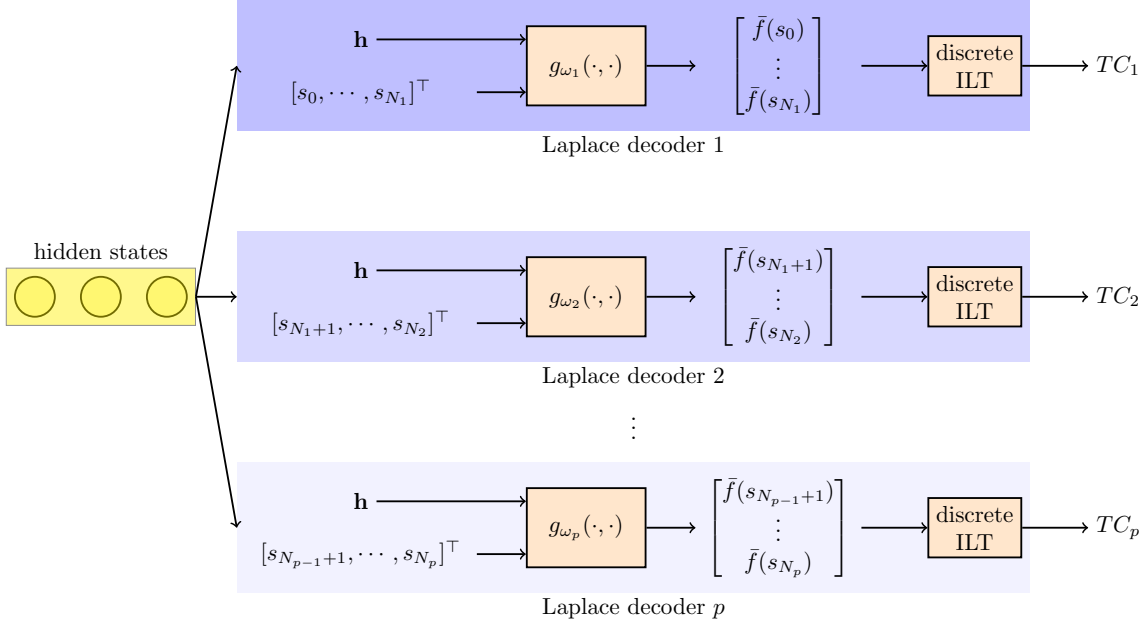


Figure 2: The details of multiple Laplace decoders

3.4. Assembler

After rebuilding a temporal function $f'(t)$ with explicit frequency information for the forecasting horizon, the assembler aims to collect the components and generate the forecasts at the desired resolutions f_r^1, \dots, f_r^m ($f_r^m \leq f_r$).

According to Nyquist-Shannon sampling theorem [29], for a given sampling resolution f_r , the maximum informative frequency that can be expressed in the original energy data is no more than $1/2f_r$, i.e., $f_c^{max} \leq 1/2f_r$. From Eq. (8), we know that the maximum informative frequency in the energy sequence data is $f_c^{max} = N/2T$. Therefore, if we reconstruct the temporal function through (6) without information loss, then the discretization parameter N can be set as:

$$N \leq 2T \frac{1}{2f_r} = Tf_r. \quad (11)$$

It tells us that by setting $N = Tf_r$, we can include all the necessary information at the resolution f_r . Also, if we need to forecast at a lower resolution, for instance $f_r^i, \forall i = 1, \dots, m$, with precise frequency information, then we can just lower N to Tf_r^i . A toy example of how the parameter N affects the resulting temporal function can be found in the appendix.

Therefore, before collecting the components, we can first set the number p and positions N_i of the split points on the frequency band according to the desired resolutions:

$$\begin{cases} p = m, \\ N_i = Tf_r^i, \forall i = 1, \dots, m. \end{cases} \quad (12)$$

Then, for the resolution f_r^i , TC_1, \dots, TC_i are the temporal components with necessary frequency

information. In assembler, a corresponding temporal function $f'_i(t)$ is formed accordingly:

$$f'_i(t) = \frac{e^{\gamma t}}{T} \left[\underbrace{\frac{\bar{f}(s_0)}{2} + \sum_{k=1}^{N_1} S_k(t)}_{TC_1} + \cdots + \underbrace{\sum_{k=N_{i-1}+1}^{N_i} S_k(t)}_{TC_i} \right]. \quad (13)$$

The final energy forecasts at the resolution f_r^i can be inferred by time steps as $\hat{\mathbf{x}}_i = [f'_i(t_{n+1}), \dots, f'_i(t_{n+K_i})]$.

4. Case studies

4.1. Data and benchmarks

We focus on multi-resolution forecasting case studies on both electricity load data [30] and wind power data [31], along with NWP data from ECMWF [32]. For each dataset, we produce forecasts at three commonly used resolutions, namely 5-minute, 15-minute, and 60-minute, for the next 24 hours.

Prevailing ML-based energy forecasting models, i.e., Multi-Layer Perceptron (MLP) and Long Short-Term Memory (LSTM), are included as benchmarks. We also include the original Neural Laplace (NL) [26] as the benchmark. It only has one Laplace decoder with a small N in discrete ILT. It should be mentioned that the curse of dimensionality may happen if we only use one decoder with a large enough N to cover the whole frequency domain. The ineffectiveness in such case can be found in the appendix. One naive model, Persistence, is included as the baseline model. The latest observed data is used as forecast, for all lead times. Besides, two widely used coordination strategies for multi-resolution forecasts [14], i.e., bottom-up-based strategy (BU) and optimized-based strategy (OPT), are also applied to each of the benchmark models. Therefore, for each resolution, we have three types of benchmark forecasts:

- Raw forecasts from: MLP, LSTM, NL, and Persistence,
- BU-based coordinated forecasts from: MLP-BU, LSTM-BU, and NL-BU,
- OPT-based coordinated forecasts from: MLP-OPT, LSTM-OPT, and NL-OPT

The BU strategy means that we only forecast on the highest resolution, and the forecasts at the lower resolutions are all generated by downsampling. The OPT strategy makes full use of forecasts at all desired resolutions, solving a generalized least squares problem with the forecasts at resolutions and the temporal structure as inputs [14].

4.2. Performance metrics

The performance metrics used for evaluating the energy forecasts include RMSE in the time domain, RMSE in the frequency domain, Total Consistency Error (TCE), and the total operation costs in the energy scheduling.

In the test set D_{test} , we denote the j th forecast at the resolution f_r^i on the following steps K_i are denoted as $\hat{\mathbf{x}}_{i,j}$. The corresponding real energy data at this resolution is denoted as $\mathbf{x}_{i,j}$. Therefore, the average RMSE in the time domain for this resolution f_r^i over the whole test set can be calculated as:

$$RMSE_{\text{time}}^i = \frac{1}{|D_{\text{test}}|} \sum_{j=1}^{|D_{\text{test}}|} \|\hat{\mathbf{x}}_{i,j} - \mathbf{x}_{i,j}\|_2, \quad (14)$$

where $|D_{\text{test}}|$ denotes the size of the test set.

For the RMSE in the frequency domain, we perform fast Fourier transform [33] (FFT) on for each test label and forecast at the highest resolution (5-minute in this paper). The resulting frequency spectrums of the j th test label and the forecast are denoted as $\mathbf{y}_{i,j}$ and $\hat{\mathbf{y}}_{i,j}$. In this way, we can compute the average RMSE in the frequency domain over the whole test set:

$$RMSE_{\text{freq}} = \frac{1}{|D_{\text{test}}|} \sum_{j=1}^{|D_{\text{test}}|} \|\hat{\mathbf{y}}_{i,j} - \mathbf{y}_{i,j}\|_2, \quad (15)$$

The unit of both RMSEs in both time and frequency domain is the same as the original energy data, i.e. kW in our case studies.

In terms of evaluating forecasting consistency, the TCE over the whole test set is calculated as follows:

$$TCE = \frac{1}{|D_{\text{test}}|} \sum_{j=1}^{|D_{\text{test}}|} \sum_{a=1}^m \sum_{b=a+1}^m \|ds(\hat{\mathbf{x}}_{b,j}, f_r^a) - \hat{\mathbf{x}}_{a,j}\|_2^2, \quad (16)$$

where the function $ds(\cdot, \cdot)$ takes in a high-resolution forecast at f_r^b and downsamples them into a low-resolution one at $f_r^a (\leq f_r^b)$. Then, the Euclidean distance of the downsampled forecasts and the original forecasts at the resolution f_r^a will be calculated. For each resolution pair (f_r^a, f_r^b) , we calculate and sum up to obtain the final TCE of the j th forecasts. The average TCE over the whole test set is then calculated as above.

As for energy scheduling, we consider an economic dispatch problem in both day-ahead and intra-day fashion, including the ramping constraints of generators, battery storage devices and so on. The details of the specific optimization problems can be found in the appendix.

4.3. Accurate forecasts at different resolutions

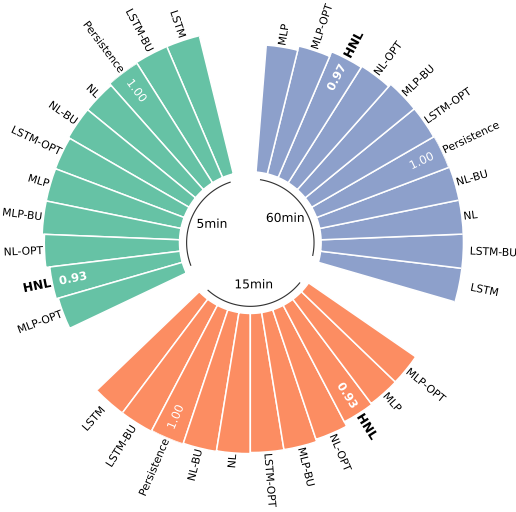
To comprehensively evaluate forecast performance, we run each forecast model 20 times with different random seeds. Then, the Root Mean Squared Error (RMSE) at each and every resolution is calculated for each seed, and the comparison of average RMSEs over all seeds is depicted in Fig. 3(a)(c).

For electricity load forecasting, the proposed HNL framework outperforms the baseline model by 7.28%, 7.08%, and 3.4%, respectively, at 5-minute, 15-minute, and 60-minute. Coordinating strategies, especially OPT-based strategy, generally lower the forecasting errors of benchmarks at all resolutions.

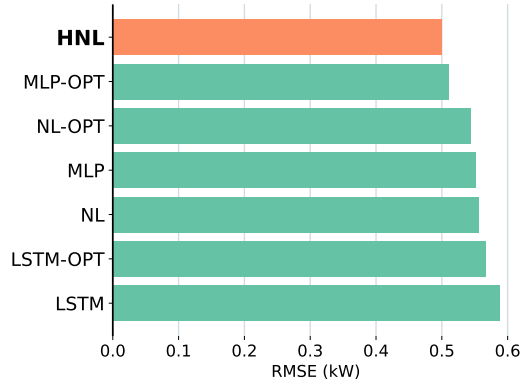
Compared to the benchmarks, HNL still shows superiority over most of them and achieves competitive performance with the state-of-the-art coordinated model (MLP-OPT). For wind power forecasting, the performance differences are more remarkable. Outperforming the baseline Persistence by at least 29%, the proposed HNL obtains the lowest forecasting error at all resolutions. It still has at least a 3% accuracy improvement over the best-performing benchmarks at each resolution. Thus, on both load and wind power forecasting, the competitive performance of HNL is obvious in terms of forecast quality in the time domain at all resolutions.

In addition to the performance in the time domain, we also show that our proposed method can capture more precise information in the frequency domain. In Fig. 3(b)(d), HNL attains the best performances in the frequency domain for both datasets, owing to the hierarchical frequency learning. With the limited capability of frequency learning, there is at least 5% performance lag for NL-based methods. Despite the close performance of MLP-OPT with HNL in the load forecasting,

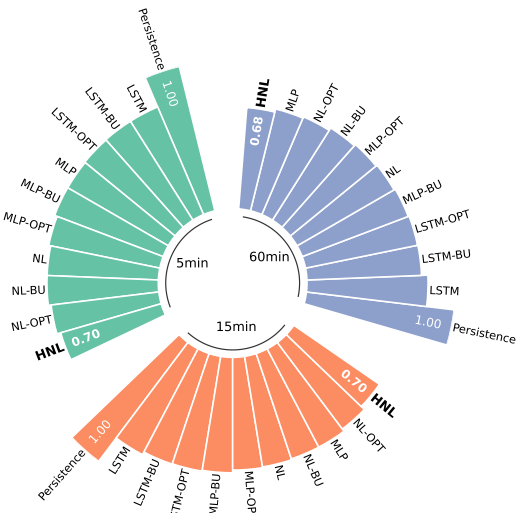
MLP-based models fail to accurately portray the actual frequency information of wind power. Similarly, mainly focusing on the temporal dependence, LSTM-based models cause around 15% more error in the frequency domain than our proposed HNL. Therefore, besides the competitive forecasting performance in the time domain, our proposed method can capture the information in the frequency domain more accurately than the prevailing forecasting models.



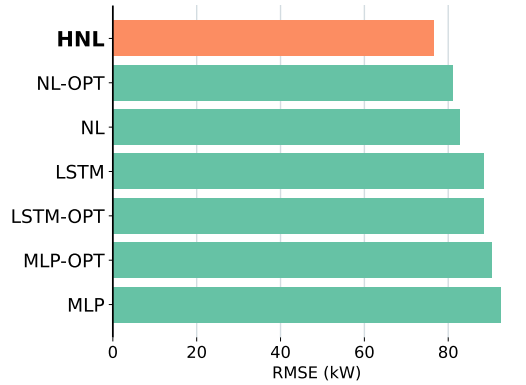
(a) Comparison on time domain (electricity load)



(b) Comparison on frequency domain (electricity load)



(c) Comparison in the time domain (wind power)



(d) Comparison in the frequency domain (wind power)

Figure 3: Comparison of forecasting errors (RMSE) in both time and frequency domain in terms of electricity load and wind power. (a) and (c) show the comparison in the time domain, while (b) and (d) show the comparison in the frequency domain at the highest resolution. For (a) and (c), RMSEs are scaled based on the performance of the Persistence model. For (b) and (d), BU-based forecasts are omitted because they are the same as the raw forecasts at the highest resolution.

4.4. Consistent forecasts across resolutions

Beyond the accuracy at each resolution, we also evaluate the consistency of the forecasts across these resolutions. Fig. 4 presents the comparison of our proposed HNL and uncoordinated models in terms of consistency. The most obvious inconsistency happens between the highest and the lowest resolution, i.e., 5-minute and 60-minute. The proposed HNL brings significant advantages in both electricity load and wind power forecasting. In contrast, the raw forecasts from the benchmarks without any coordination apparently cause extremely high consistency errors.

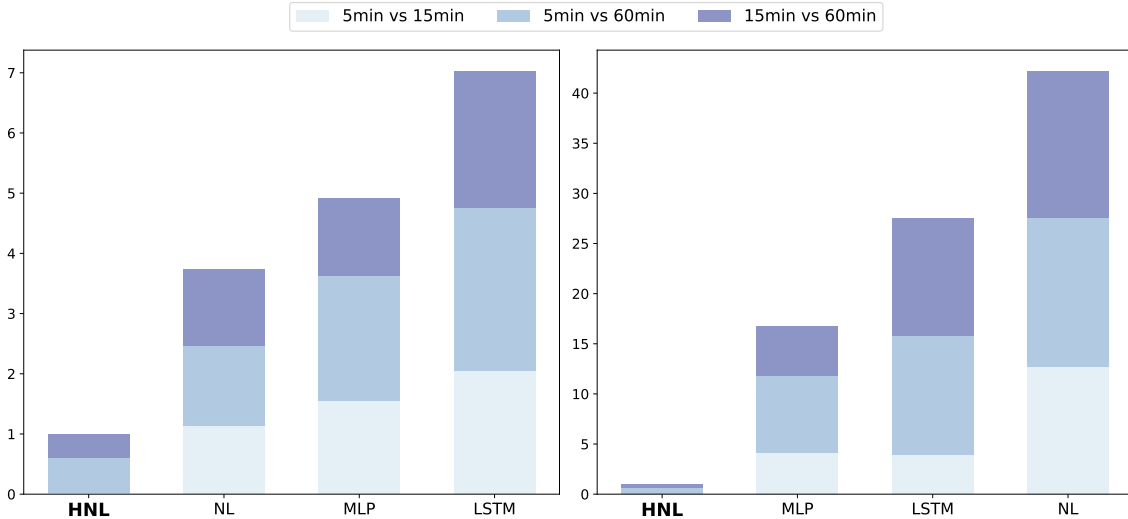


Figure 4: TCE comparison. The TCEs are scaled based on the performance of the HNL. Left: electricity load forecasts. Right: wind power forecasts

Besides, the leading position of HNL in wind power forecasting is more striking than electricity load forecasting. Our proposed HNL is at least 10 times better than the benchmarks in the wind power forecasting case. Fig. 5 depicts an example of multi-resolution forecasts comparison for wind power generation.

From the perspective of accuracy, traditional deep learning-based models, LSTM and MLP, fail to produce accurate forecasts in terms of both time and frequency domains. Especially in high-resolution situations, forecasts from MLP-related models introduce redundant high-frequency information. Though the forecasts from NL-related models track the trend of wind power, they fail to represent the information across relevant frequency ranges, owing to the limited frequency learning capability. These phenomena reflect the conclusion that the HNL method yield more accurate forecasts when assessed in both time and frequency domains.

From the perspective of consistency, there appears abrupt gaps during the transition of forecasting resolutions among raw forecasts from benchmarks. For example, forecasts from LSTM somehow become lower if the forecasting resolution is switched from 15-minute to 60-minute. Both post-coordination strategies (BU and OPT) eliminate these jumps in the raw forecasts, enhancing the consistency of the benchmark approaches. In contrast, even without any coordination strategies, the forecasts from HNL can still behave consistently across resolutions. The transition from

one resolution to another is natural because of the framework design that generates energy forecasts at different resolutions hierarchically. Therefore, beyond the high accuracy in both time and frequency domains, forecasts from HNL are highly consistent across resolutions.

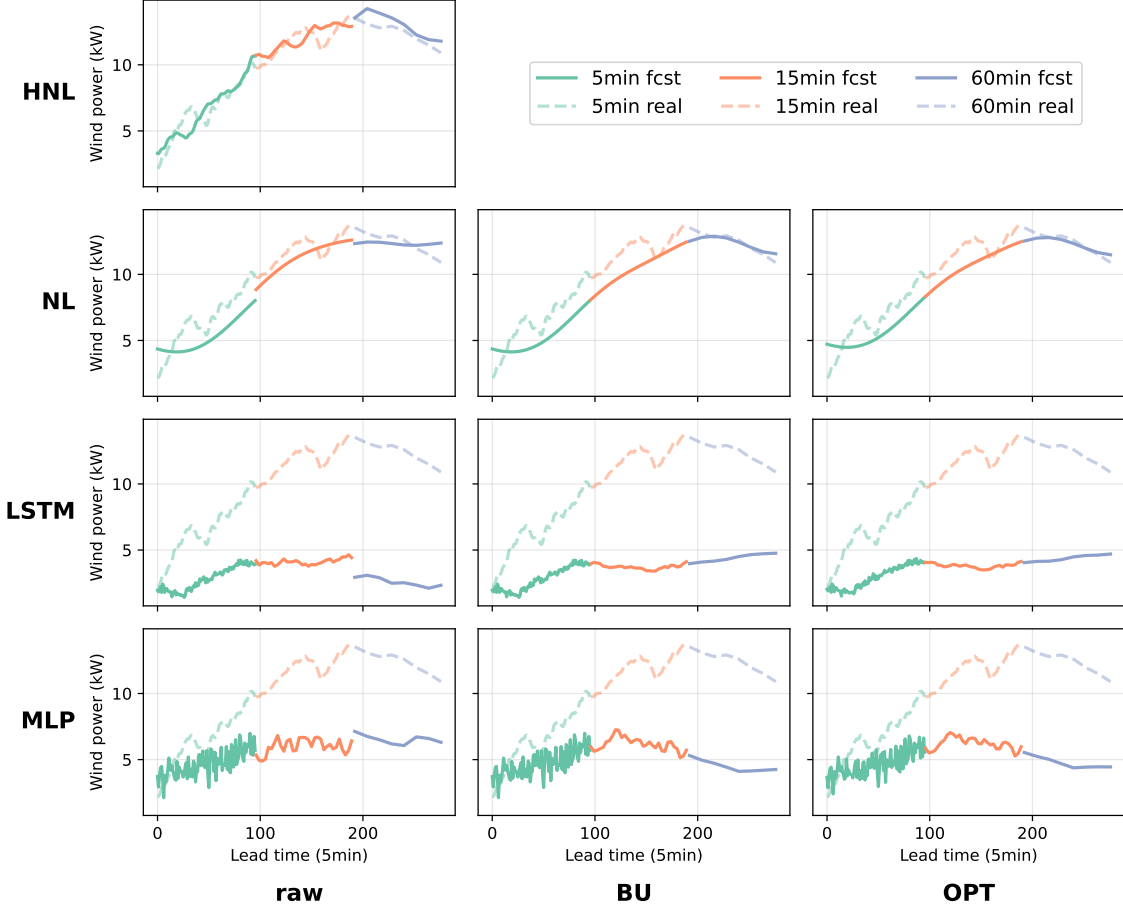


Figure 5: Multi-resolution forecasts comparison on wind power. One forecasting day is separated into three even horizons for demonstration. Each row represents the model type, and each column represents the coordination strategy. The proposed HNL belongs to raw forecasting methods since it doesn't experience any coordination.

4.5. Supportive effects for decision making

The final goal of energy forecasting is to support decision-making, for a range of operational problems. Besides the accuracy and consistency of the forecasts, we also analyze the operational benefits brought by the forecasts in the downstream decision-making process.

Specifically, we start with the day-ahead energy scheduling problem for a local region. It is a cost minimization problem that takes the 24-hour load forecast as the demand to be met, and outputs the optimal schedule of energy supply for each time step for the next day. Though the scheduling horizon is fixed to 24 hours, the scheduling resolution can vary, depending on how detailed the

schedules are expected. Consequently, for different scheduling resolutions, the forecasts with the corresponding resolutions are required. Since it is impossible to forecast perfectly, additional real-time operations like external energy purchases are needed to balance the real energy demand and scheduled energy supply. This will cause extra operation costs. Therefore, regarding the extra costs, we evaluate the operational effects of the day-ahead forecasts at each resolution. The details of the settings of the day-ahead energy scheduling problem can be found in the appendix.

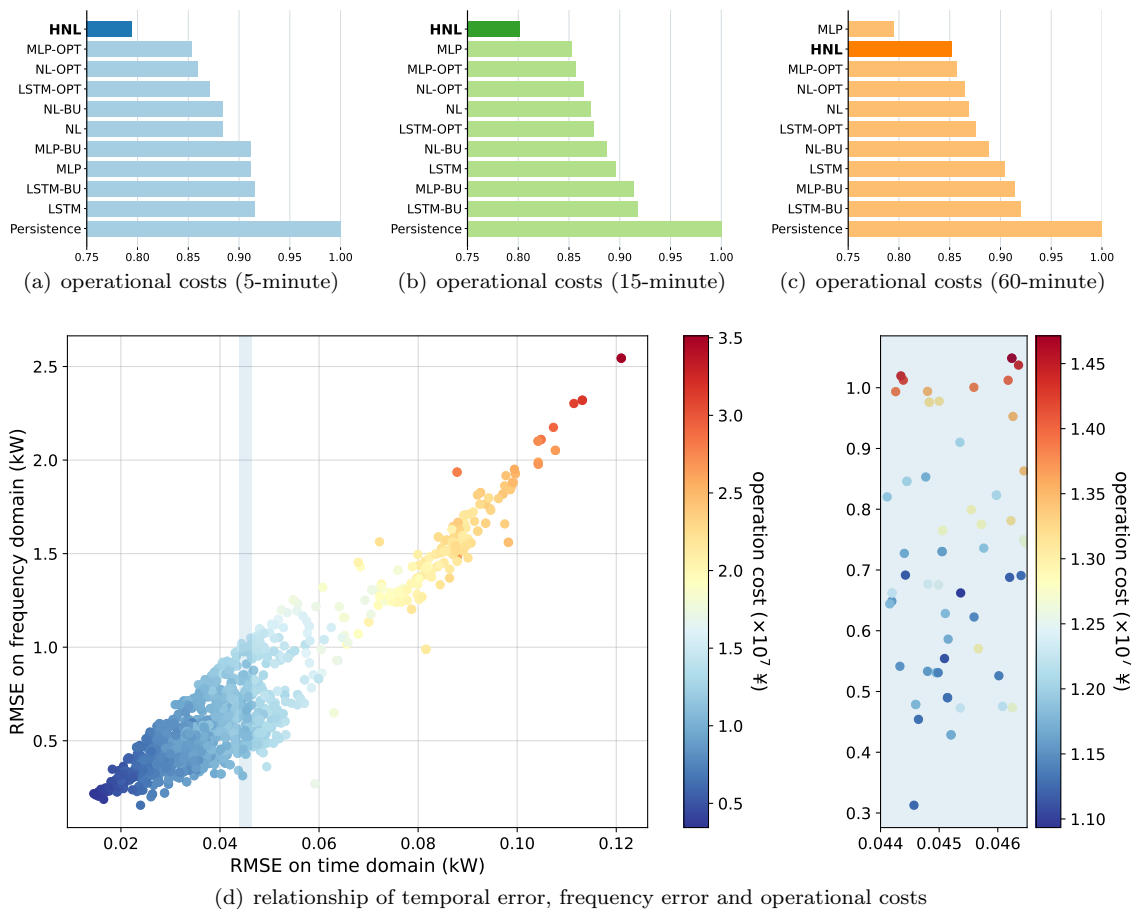


Figure 6: Comparison of total operational costs in day-ahead scheduling. (a),(b) and (c) show the comparison based on 5-minute, 15-minute, and 60-minute resolution forecasts, respectively. The operational costs in (a), (b), and (c) are scaled based on the performance of the Persistence model. (d) shows the relationship among temporal error, frequency error, and operational costs at the highest resolution.

Fig. 6(a)(b)(c) illustrate the comparison of operational costs from all forecasting models at different resolutions. We observe that with the resolution becoming higher, the advantage of our proposed HNL is more remarkable. Compared to the baseline model, the percentage savings of the operational costs are 14.8%, 19.9%, and 20.6% at the resolutions of 60-minute, 15-minute, and 5-minute, respectively. For the 60-minute resolution, the prevailing machine-learning model MLP

even achieves lower operation costs. The possible reason is that hourly day-ahead forecasts only require 24 prediction points. The number of output dimensions is relatively small and contains limited frequency information. Traditional machine learning methods can perform well in such low-dimension forecasting settings.

However, when it comes to the higher resolution cases, the energy curves will reveal more detailed frequency information. It may also affect the operation results because of more strict ramping constraints and frequent battery charging and discharging operations. We collected each day-ahead forecast from all the models and computed the corresponding temporal RMSE, frequency RMSE, and total operational cost. Fig. 6 (d) depicts that forecasting errors in both time and frequency domains generally follow a linear relationship. Lower errors jointly in the time and frequency domains lead to lower additional operation costs. The blue shaded part shows a large proportion of special situations. The temporal errors of the forecasts are at a similar level, but the frequency errors varies a lot. In this vertical slice, we observe that with similar temporal accuracy, generally, the lower the errors in the frequency domain are, the lower the additional operational costs are in the day-ahead scheduling. Therefore, while our proposed HNL approach achieved similar temporal performance with the state-of-the-art benchmarks, the performance advantage of HNL in the frequency domain makes it substantially better in supporting decision-making.

In addition to day-ahead scheduling, energy system operators launch intra-day scheduling on a regular basis (for example, every 4 hours) to modify the day-ahead schedules. Distinct from day-ahead scheduling, intra-day scheduling concentrates on the operation decisions for a shorter horizon (usually 4 hours) and higher resolution (at least 15-minute). This is also a cost minimization problem which requires the updated energy forecasts at the according horizon and resolution.

Here, to conduct a more realistic case study, we jointly consider the day-ahead and intra-day scheduling, and denote as integrated scheduling. Apparently, in this case, multi-resolution energy forecasts cooperatively contribute to the final operation decisions. To quantify the value of multi-resolution forecasts in this context, we calculate the total operation costs, including the day-ahead costs for energy arrangements, the intra-day costs for adjustments, and the real-time costs for energy balancing. Details on integrated scheduling are attached in the appendix.

To fully illustrate the operational effects of both energy demand and supply forecasts, we enumerate all the possible pairs of forecasting approaches for electricity load and wind power generation, and calculate the resulting total operational costs. Fig. 7 presents the results in integrated scheduling of different model pairs. It is proved that our proposed HNL brings significant supportive effects in this large-scale realistic integrated scheduling case. Compared with the Persistence-based methods, the saved costs are more than 25%. We also observe that, in this case, the differences between each column were remarkable, which means the wind forecasting methods dominate in the integrated scheduling. This is because the penetration rate of the wind power is set as a relatively high ratio (50%). The results with other wind power penetration rates can be found in the appendix.

Therefore, with high accuracy in both the time and frequency domain, our proposed energy forecasting framework, HNL, can provide more supportive forecasts for day-ahead scheduling. The relationship between temporal error, frequency error, and operation costs is explored as well. The striking advantage of HNL continues to be effective in the integrated scheduling process.

5. Conclusion and future works

In this paper, an innovative energy forecasting framework, HNL, is proposed for multi-resolution forecasting, providing unified modeling at different resolutions. Without the need for multiple mod-

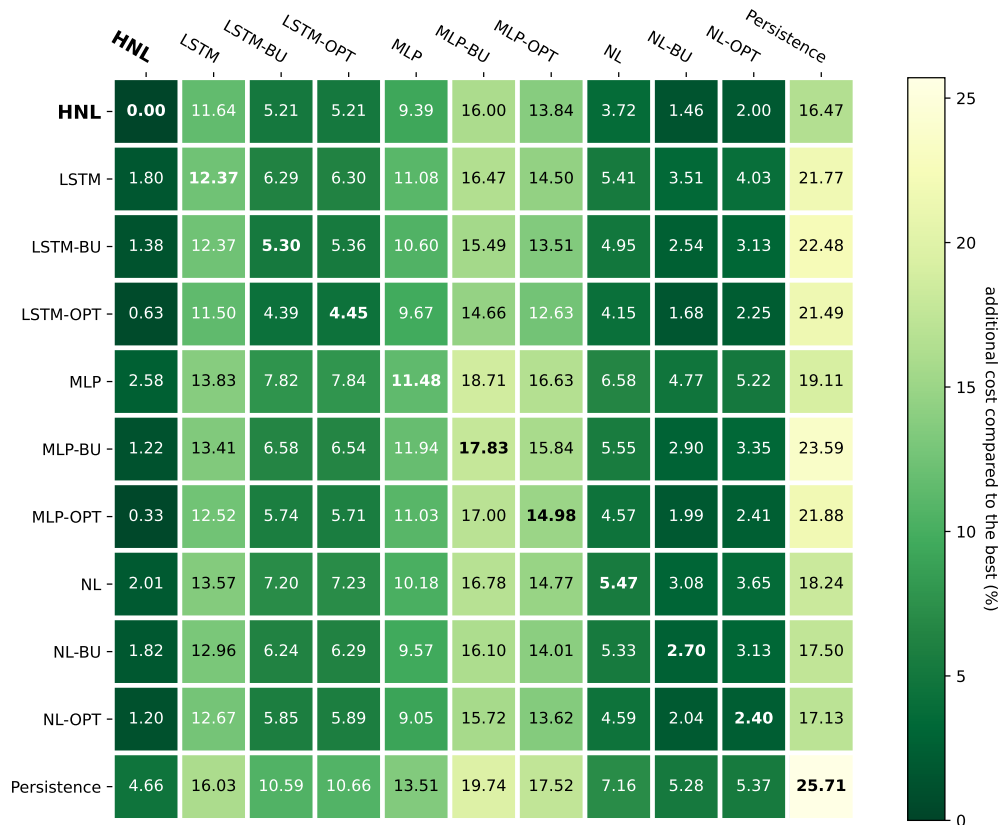


Figure 7: Comparison of operational costs for integrated scheduling. Columns: wind forecasting methods. Rows: electricity load forecasting methods. Entries: additional costs compared to the best in percentage value. Zero means the most economic situation.

els, the proposed framework is designed to characterize the hierarchical frequency information of the energy data based on the Laplace transform. Forecasts at desired resolutions can be flexibly produced by aggregating corresponding temporal components based on the Shannon sampling theorem.

Case studies have demonstrated that the HNL framework attains satisfactory accuracy improvement on the competitive benchmarks and baseline model. Thanks to the hierarchical framework design, more seamless forecasts across the resolutions can be generated than benchmarks. The focus on frequency learning gives the HNL an additional advantage in the accuracy in the frequency domain where the HNL can capture more precise variation patterns of the energy data. More importantly, in the follow-up day-ahead energy scheduling, this advantage of HNL can lead to better schedules with lower operation costs. The analysis of the relationship between temporal error, frequency error, and operation costs further implies that with similar temporal accuracy, forecasts with more precise frequency information may lead to better decisions. Consequently, it is suggested that accuracy in both time and frequency domains matters in energy forecasting, especially for the

downstream decision-making. Besides, in the complicated integrated scheduling, multi-resolution forecasts of electricity load and wind power from HNL resulted in a salient reduction in terms of the total operation costs as well.

Though the HNL framework has achieved a satisfying performance on energy forecasting, there are still some aspects that deserve to be further explored. First, when it comes to a very high-resolution situation, there exists uncertain patterns in the energy curves, especially in the context of renewable energies [34, 35], which is the nature of the energy data. Therefore, it is necessary to extend the current HNL framework into a probabilistic one to model the uncertainty of the forecasts. Besides, Laplace Transform can only tell what kind of temporal components exist in the data but can not tell when and what components are dominant. We will further inspect how to treat other time-frequency transformations as decoders in energy forecasting. In this way, we can analyze the forecast results and their operational effects for specific time periods (for example, peak hours) in a refined manner.

Nevertheless, the proposed framework obtains fair performance over competitive benchmarks, which provides a unified perspective to deal with multi-resolution energy forecasting by capturing the hierarchical frequency information in the energy data. In addition, forecasts from the proposed framework are proven to have prominent supportive effects for energy scheduling, which fulfill the research gap of multi-resolution energy forecasting and its resulting operation effects.

6. Data availability

The load and wind power datasets are available respectively at <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/X9MIDJ> and <https://www.nrel.gov/grid/wind-toolkit.html>. The NWP data is available from the ECMWF website <https://www.ecmwf.int/en/forecasts/dataset/operational-archive> after registration as real-name users (for researchers based in Europe, at least). The above mentioned datasets have been processed and deposited in our GitHub repository <https://github.com/WillWang1113/MultiresolutionForecasting>.

7. Code availability

The code and data for the experiments are publicly available on <https://github.com/WillWang1113/MultiresolutionForecasting>.

References

- [1] Stephen Haben, Siddharth Arora, Georgios Giasemidis, Marcus Voss, and Danica Vukadinović Greetham. Review of low voltage load forecasting: Methods, applications, and recommendations. *Applied Energy*, 304:117798, 2021.
- [2] Yun Wang, Runmin Zou, Fang Liu, Lingjun Zhang, and Qianyi Liu. A review of wind speed and wind power forecasting with deep neural networks. *Applied Energy*, 304:117766, 2021.
- [3] Tao Hong, Pierre Pinson, Yi Wang, Rafał Weron, Dazhi Yang, and Hamidreza Zareipour. Energy forecasting: A review and outlook. *IEEE Open Access Journal of Power and Energy*, 7:376–388, 2020.

- [4] Benjamin F Hobbs, Suradet Jitprapaikularn, Sreenivas Konda, Vira Chankong, Kenneth A Loparo, and Dominic J Maratukulam. Analysis of the value for unit commitment of improved load forecasts. *IEEE Transactions on Power Systems*, 14(4):1342–1348, 1999.
- [5] R Tawn and J Browell. A review of very short-term wind and solar power forecasting. *Renewable and Sustainable Energy Reviews*, 153:111758, 2022.
- [6] Daniel Vázquez Pombo, Tuhfe Göçmen, Kaushik Das, and Poul Sørensen. Multi-horizon data-driven wind power forecast: From nowcast to 2 days-ahead. In *2021 International Conference on Smart Energy Systems and Technologies (SEST)*, pages 1–6. IEEE, 2021.
- [7] L Suganthi and Anand A Samuel. Energy models for demand forecasting—a review. *Renewable and sustainable energy reviews*, 16(2):1223–1240, 2012.
- [8] Huaizhi Wang, Zhenxing Lei, Xian Zhang, Bin Zhou, and Jianchun Peng. A review of deep learning for renewable energy forecasting. *Energy Conversion and Management*, 198:111799, 2019.
- [9] Fotios Petropoulos, Daniele Apiletti, Vassilios Assimakopoulos, Mohamed Zied Babai, Devon K Barrow, Souhaib Ben Taieb, Christoph Bergmeir, Ricardo J Bessa, Jakub Bijak, John E Boylan, et al. Forecasting: theory and practice. *International Journal of Forecasting*, 38(3):705–871, 2022.
- [10] Georges Kariniotakis. *Renewable energy forecasting: from models to applications*. Woodhead Publishing, 2017.
- [11] Han Li, Zhe Wang, Tianzhen Hong, Andrew Parker, and Monica Neukomm. Characterizing patterns and variability of building electric load profiles in time and frequency domains. *Applied Energy*, 291:116721, 2021.
- [12] Yvenn Amara-Ouali, Matteo Fasiolo, Yannig Goude, and Hui Yan. Daily peak electrical load forecasting with a multi-resolution approach. *International Journal of Forecasting*, 39(3):1272–1286, 2023.
- [13] Hui Liu, Zhu Duan, and Chao Chen. Wind speed big data forecasting using time-variant multi-resolution ensemble model with clustering auto-encoder. *Applied Energy*, 280:115975, 2020.
- [14] George Athanasopoulos, Rob J Hyndman, Nikolaos Kourentzes, and Fotios Petropoulos. Forecasting with temporal hierarchies. *European Journal of Operational Research*, 262(1):60–74, 2017.
- [15] Dazhi Yang, Hao Quan, Vahid R Disfani, and Carlos D Rodríguez-Gallegos. Reconciling solar forecasts: Temporal hierarchy. *Solar Energy*, 158:332–346, 2017.
- [16] Peter Nystrup, Erik Lindström, Pierre Pinson, and Henrik Madsen. Temporal hierarchies with autocorrelation for load forecasting. *European Journal of Operational Research*, 280(3):876–888, 2020.
- [17] Maryam Nejati, Nima Amjady, and Hamidreza Zareipour. A new multi-resolution closed-loop wind power forecasting method. *IEEE Transactions on Sustainable Energy*, 2023.

- [18] Hui Liu, Rui Yang, and Zhu Duan. Wind speed forecasting using a new multi-factor fusion and multi-resolution ensemble model with real-time decomposition and adaptive error correction. *Energy Conversion and Management*, 217:112995, 2020.
- [19] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.
- [20] Ramin Hasani, Mathias Lechner, Alexander Amini, Lucas Liebenwein, Aaron Ray, Max Tschaikowski, Gerald Teschl, and Daniela Rus. Closed-form continuous-time neural networks. *Nature Machine Intelligence*, 4(11):992–1003, 2022.
- [21] Patrick Kidger, James Morrill, James Foster, and Terry Lyons. Neural controlled differential equations for irregular time series. *Advances in Neural Information Processing Systems*, 33:6696–6707, 2020.
- [22] Mehrnaz Anvari, Elisavet Proedrou, Benjamin Schäfer, Christian Beck, Holger Kantz, and Marc Timme. Data-driven load profiles and the dynamics of residential electricity consumption. *Nature communications*, 13(1):4593, 2022.
- [23] Changfei Zhao, Can Wan, and Yonghua Song. Cost-oriented prediction intervals: On bridging the gap between forecasting and decision. *IEEE Transactions on Power Systems*, 37(4):3048–3062, 2021.
- [24] Jialun Zhang, Yi Wang, and Gabriela Hug. Cost-oriented load forecasting. *Electric Power Systems Research*, 205:107723, 2022.
- [25] Guo Li and Hsiao-Dong Chiang. Toward cost-oriented forecasting of wind power generation. *IEEE Transactions on Smart Grid*, 9(4):2508–2517, 2016.
- [26] Samuel I Holt, Zhaozhi Qian, and Mihaela van der Schaar. Neural laplace: Learning diverse classes of differential equations in the laplace domain. In *International Conference on Machine Learning*, pages 8811–8832. PMLR, 2022.
- [27] Kenny S Crump. Numerical inversion of laplace transforms using a fourier series approximation. *Journal of the ACM (JACM)*, 23(1):89–96, 1976.
- [28] Kristopher L Kuhlman. Review of inverse laplace transform algorithms for laplace-space numerical approaches. *Numerical Algorithms*, 63:339–355, 2013.
- [29] Claude E Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.
- [30] Christoph J Meinrenken, Noah Rauschkolb, Sanjmeet Abrol, Tuhin Chakrabarty, Victor C Decalf, Christopher Hidey, Kathleen McKeown, Ali Mehmani, Vijay Modi, and Patricia J Culligan. Mfred, 10 second interval real and reactive power for groups of 390 us apartments of varying size and vintage. *Scientific Data*, 7(1):375, 2020.
- [31] Caroline Draxl, Andrew Clifton, Bri-Mathias Hodge, and Jim McCaa. The wind integration national dataset (wind) toolkit. *Applied Energy*, 151:355–366, 2015.
- [32] MARS - the ECMWF meteorological archive, 2018. Pages: 1 hour Publication Title: Software and computing services.

- [33] E Oran Brigham. *The fast Fourier transform and its applications*. Prentice-Hall, Inc., 1988.
- [34] Jianxiao Wang, Liudong Chen, Zhenfei Tan, Ershun Du, Nian Liu, Jing Ma, Mingyang Sun, Canbing Li, Jie Song, Xi Lu, et al. Inherent spatiotemporal uncertainty of renewable power in china. *Nature Communications*, 14(1):5379, 2023.
- [35] Lynn H Kaack, Jay Apt, M Granger Morgan, and Patrick McSharry. Empirical prediction intervals improve energy forecasting. *Proceedings of the National Academy of Sciences*, 114(33):8752–8757, 2017.

Appendix A Energy scheduling

A.1 Day-ahead scheduling

$$\min_{\substack{\{P_{ji}, P_i^c, \\ P_i^d, SOC_i, \phi_i\}}} C_g + C_b \quad (17)$$

$$s.t. \quad C_g = \sum_{i=1}^N \sum_{j=1}^J (a_j P_{ji}^2 + b_j P_{ji} + c_j) \quad (18)$$

$$C_b = \sum_{i=1}^N \pi \cdot (\eta_c P_i^c + P_i^d / \eta_d) \quad (19)$$

$$\sum_{j=1}^J P_{ji} + P_i^d = \widehat{L}_i + P_i^c \quad : \forall i, \quad (20)$$

$$0 \leq P_{ji} \leq \overline{P}_j \quad : \forall j, \forall i, \quad (21)$$

$$P_{ji} - P_{j(i-1)} \leq R_j^U \cdot \Delta t \quad : \forall j, \forall i \geq 2, \quad (22)$$

$$P_{j(i-1)} - P_{ji} \leq R_j^D \cdot \Delta t \quad : \forall j, \forall i \geq 2, \quad (23)$$

$$Cap \cdot SOC_i = Cap \cdot SOC_{\text{init}} + (P_i^c \cdot \eta_c - P_i^d / \eta_d) \cdot \Delta t \quad : i = 1, \quad (24)$$

$$Cap \cdot SOC_i = Cap \cdot SOC_{i-1} + (P_i^c \cdot \eta_c - P_i^d / \eta_d) \cdot \Delta t \quad : \forall i \geq 2, \quad (25)$$

$$\underline{SOC} \leq SOC_i \leq \overline{SOC} \quad : \forall i, \quad (26)$$

$$0 \leq P_i^c \leq P_j^{\text{batt}} \cdot \phi_i \quad : \forall i, \quad (27)$$

$$0 \leq P_i^d \leq P_j^{\text{batt}} \cdot (1 - \phi_i) \quad : \forall i, \quad (28)$$

$$\phi_i \in \{0, 1\}, \quad : \forall i. \quad (29)$$

In the day-ahead scheduling problem, on the basis of power balance (Eq. (20)), we consider the generators with ramping constraints (Eq. (21) - Eq. (23)) for the basic supply and the battery storage (Eq. (24) - Eq. (29)) for flexible supply. i is the index for time unit and j is the index for the generators. P_{ji} represents the power output of the j th generator at i th time unit. P_i^c and P_i^d represent the charging and discharging power at i th time unit. ϕ_i is a binary variable which controls charging or discharging. SOC_i is the state of charging of the battery. The power flow constraints are omitted for simplicity.

Given the load forecasts $\widehat{\mathbf{L}} = [\widehat{L}_1, \dots, \widehat{L}_N]$, the aim of the day-ahead scheduling is to give the day-ahead schedule \mathbf{P} that minimizes the sum of the costs of the generators C_g and the degradation costs of the battery C_b . Further, we denote this day-ahead operation costs and the day-ahead schedule as a function of forecasts, i.e. $C_{da}(\widehat{\mathbf{L}}), \mathbf{P}(\widehat{\mathbf{L}})$, and the operation costs based on the perfect forecasts as $C_{da}^* = C_{da}(\mathbf{L}), \mathbf{P}^* = \mathbf{P}(\mathbf{L})$.

Based on the day-ahead schedule of generators P_{ji}^{da} , the battery storage schedule will be re-adjusted to balance the supply and demand according to the real load L . This forms the following problem:

$$\min_{\{P_i^c, P_i^d, SOC_i, \phi_i\}} C_g + C_b \quad (30)$$

$$s.t. \quad (18) - (19) \quad (31)$$

$$\sum_{j=1}^J P_{ji} + P_i^d = L_i + P_i^c \quad : \forall i, \quad (32)$$

$$P_{ji} = P_{ji}^{da} \quad : \forall j, \forall i, \quad (33)$$

$$(24) - (29) \quad (34)$$

Therefore, we can see that the costs of generators C_g is not changing (Eq. (33)) while the costs of the battery C_b will be re-calculated according to the difference between the scheduled supply and the real load. We also denote this real-time operation costs of day-ahead schedule as a function of day-ahead schedules $C_{rt}(\mathbf{P}(\widehat{\mathbf{L}}))$ and for the perfect forecasts, the corresponding cost is $C_{rt}^* = C_{rt}(\mathbf{P}^*)$. Consequently, in the perfect forecasting situation, we don't have to re-adjust the battery schedule and have $C_{rt}^* = C_{da}^*$. However, for the imperfect forecasts, we will have $C_{rt}(\mathbf{P}(\widehat{\mathbf{L}})) > C_{da}(\widehat{\mathbf{L}})$. To evaluate the operational effects of the forecasts caused by imperfection, we calculate the additional operation costs compared to the perfect situation and obtain $C_{rt} - C_{da}^*$.

A.2 Integrated scheduling

$$\begin{aligned} & \min_{P_{ji}, W_i, W_i^c} (C_g + C_w) \cdot \Delta t \\ & s.t. \quad C_w = WC \cdot W_i^c, \\ & \quad W_i + \sum_{j=1}^J P_{ji} = \widehat{L}_i^{low} \quad : \forall i, \\ & \quad W_i + W_i^c = \widehat{W}_i^{low} \quad : \forall i, \\ & \quad 0 \leq W_i \leq \widehat{W}_i^{low} \quad : \forall i, \\ & \quad 0 \leq W_i^c \leq \widehat{W}_i^{low} \quad : \forall i, \\ & \quad (18), (21) - (23) \end{aligned}$$

In the large-scale integrated scheduling, we first considered a day-ahead scheduling problem as illustrated above. Here, considering the difference of scheduling resolution between day-ahead and intra-day scheduling, the objective function is scaled by the time interval.

We also include the wind power into the consideration. W_i represents the wind power that is used as power supply while W_i^c represents the wind power that is curtailed. The curtailed wind

power will be punished as C_w . Beyond the load forecasts \widehat{L}_i^{low} , the wind forecasts \widehat{W}_i^{low} are also needed. Both of them are required at a relatively low resolution for day-ahead scheduling, i.e. 60min level. The aim is to determine the schedules of the generators P_{ji}^{da} .

$$\begin{aligned}
& \min_{\substack{\{W_i, W_i^c, P_i^c, \\ P_i^d, SOC_i, \phi_i\}}} (C_g + C_w + C_b) \cdot \Delta t \\
s.t. \quad & W_i + P_i^c + \sum_{j=1}^J P_{ji} = \widehat{L}_i^{high} + P_i^d \quad : \forall i, \\
& P_{ji} = P_{ji}^{da} \quad : \forall i, \forall j, \\
& W_i + W_i^c = \widehat{W}_i^{high} \quad : \forall i, \\
& 0 \leq W_i \leq \widehat{W}_i^{high} \quad : \forall i, \\
& 0 \leq W_i^c \leq \widehat{W}_i^{high} \quad : \forall i, \\
& (24) - (29).
\end{aligned}$$

On the basis of the day-ahead schedule of generators P_{ji}^{da} , the battery storage schedules P_i^c, P_i^d will be considered and determined to balance the supply and demand according to the more timely load and wind power forecasts \widehat{L}_i^{high} and \widehat{W}_i^{high} . This intra-day scheduling will be run every 4 hours at a high resolution, i.e. 5min level, for the next 4 hours.

$$\begin{aligned}
& \min_{\{W_i, W_i^c, P_i^p, P_i^n\}} (C_g + C_w + C_b + \sum_{i=1}^N (P_i^p \cdot \pi_p + P_i^n \cdot \pi_n)) \cdot \Delta t \\
s.t. \quad & W_i + P_i^c + P_i^p + \sum_{j=1}^J P_{ji} = L_i^{high} + P_i^d + P_i^n \quad : \forall i, \\
& P_{ji} = P_{ji}^{da} \quad : \forall i, \forall j, \\
& P_i^c = P_i^{id,c} \quad : \forall i, \\
& P_i^d = P_i^{id,d} \quad : \forall i, \\
& W_i + W_i^c = W_i^{high} \quad : \forall i, \\
& 0 \leq W_i \leq W_i^{high} \quad : \forall i, \\
& 0 \leq W_i^c \leq W_i^{high} \quad : \forall i, \\
& (24) - (29).
\end{aligned}$$

Finally, given the schedules of generators P_{ji}^{da} , the battery $P_i^{id,c}, P_i^{id,d}$ and the real-time load and wind power L^{high}, W^{high} , the real-time positive/negative imbalances P_i^p, P_i^n will be unveiled, and the corresponding operations that compensate the imbalances will cause extra costs $P_i^p \cdot \pi_p + P_i^n \cdot \pi_n$. Consequently, the total costs of the integrated scheduling of this day can be obtained as the objective function shows in the above problem.

Appendix B Experimental results

B.1 Toy example

Here, a toy example is shown to demonstrate the method to control the frequency we want to learn through Laplace decoder. The following figure shows the effect of frequency parameter N .

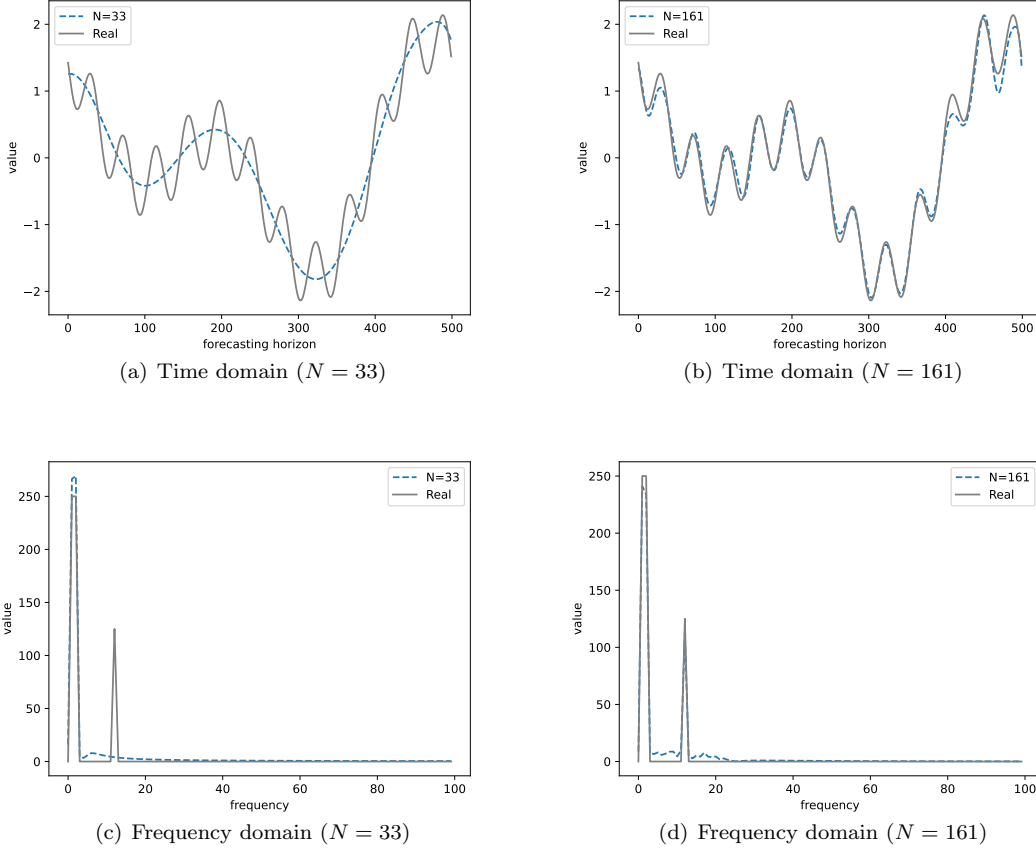


Figure 8: A toy example of the effect of frequency parameter N . The data follows $f(t) = \sin t + \sin 2t + 0.5 \sin 12t$. (a) and (b) show the forecasting results on the time domain while (c) and (d) show the corresponding spectrum on frequency domain. The forecasts only show the low frequency information with $N = 33$, but if we set $N = 161$, which is a relatively large value, the forecasts are able to capture the high frequency information.

B.2 Ineffectiveness of single Laplace decoder

We launched the experiment with one large Laplace decoder, i.e. Neural Laplace with a large enough frequency parameter N , trying to learn the whole spectrum of the energy data. However, the experiment turns out to be ineffective when only using one large Laplace decoder. The forecasting result can be observed in Fig. 9.

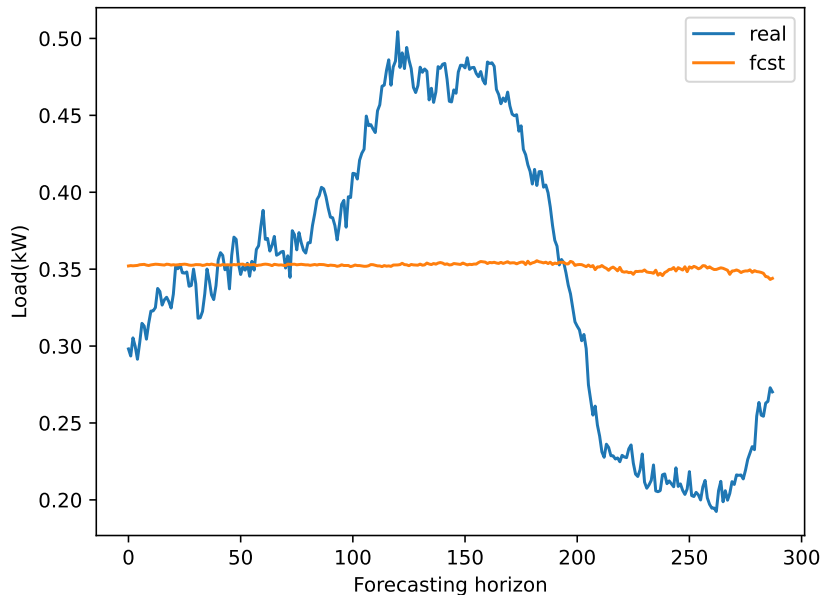


Figure 9: Ineffectiveness when only using one large Laplace decoder

Therefore, when launching the benchmarks, we set the frequency parameter in the Neural Laplace as $N = 33$ which is the same setting in the original code. In comparison, our proposed Hierarchical Laplace use multiple Laplace decoders to learn the frequency information in the energy data.

B.3 Numerical experiment results

We only show the visualization of the experiment results in the main text. This subsection will show the original numeric results of all the bar charts in the main text. It should be mentioned that the magnitudes of the load data and wind data differ a lot, therefore, the following results on the load dataset are scaled to demonstrate clearly. Related tables include Table. 1-5.

B.4 Operation costs comparison under different penetration rates

In the large-scale integrated scheduling problem, the penetration rate of wind power will determine which kind of forecasts (load or wind power) that plays a more critical role. Fig. 10 and Fig. 11 show the results under different penetration rate of wind power.

It can be seen that with the penetration rate get higher, the more distinction is shown among columns. This implies that the wind power forecasting methods gradually dominates the operation results when the penetration rate is higher.

When the penetration rate is low (0.2), where load forecasting dominates, the lowest operation costs are achieved in the row (load forecasting methods) of HNL. Similarly, when the penetration rate is high (0.8), where wind power forecasting dominates, the lowest operation costs happens

Table 1: RMSE comparison on load dataset ($\times 10^{-2}$)

Methods	5min	15min	60min
LSTM	3.751 ± 0.142	3.809 ± 0.192	3.746 ± 0.159
LSTM-BU	3.751 ± 0.142	3.708 ± 0.144	3.622 ± 0.146
LSTM-OPT	3.597 ± 0.099	3.552 ± 0.100	3.463 ± 0.102
MLP	3.591 ± 0.067	3.331 ± 0.114	3.130 ± 0.040
MLP-BU	3.591 ± 0.067	3.547 ± 0.069	3.462 ± 0.071
MLP-OPT	3.372 ± 0.060	3.325 ± 0.061	3.236 ± 0.062
NL	3.638 ± 0.129	3.579 ± 0.164	3.503 ± 0.150
NL-BU	3.638 ± 0.129	3.595 ± 0.131	3.501 ± 0.132
NL-OPT	3.531 ± 0.099	3.486 ± 0.100	3.390 ± 0.101
Persistence	3.697 ± 0.000	3.615 ± 0.000	3.479 ± 0.000
HNL	3.428 ± 0.082	3.359 ± 0.079	3.360 ± 0.068

Table 2: RMSE comparison on wind power dataset

Methods	5min	15min	60min
LSTM	5.189 ± 0.097	5.308 ± 0.196	5.403 ± 0.160
LSTM-BU	5.189 ± 0.097	5.177 ± 0.097	5.118 ± 0.098
LSTM-OPT	5.174 ± 0.086	5.162 ± 0.086	5.103 ± 0.087
MLP	5.149 ± 0.091	4.928 ± 0.119	4.680 ± 0.087
MLP-BU	5.149 ± 0.091	5.126 ± 0.090	5.063 ± 0.090
MLP-OPT	5.028 ± 0.075	5.004 ± 0.073	4.941 ± 0.074
NL	4.937 ± 0.380	4.972 ± 0.267	4.999 ± 0.349
NL-BU	4.937 ± 0.380	4.930 ± 0.381	4.869 ± 0.384
NL-OPT	4.813 ± 0.312	4.805 ± 0.312	4.743 ± 0.316
Persistence	6.596 ± 0.000	6.604 ± 0.000	6.640 ± 0.000
HNL	4.642 ± 0.252	4.617 ± 0.248	4.538 ± 0.251

in the column (wind power forecasting methods) of HNL. It suggested that HNL shows dominant operational effects in all situations considered.

Table 3: RMSE comparison on frequency domain

Method	Load ($\times 10^{-2}$)	Wind power
LSTM	58.771 \pm 3.180	88.487 \pm 2.71
LSTM-OPT	56.694 \pm 2.297	88.548 \pm 2.412
MLP	55.202 \pm 1.805	92.625 \pm 2.526
MLP-OPT	51.102 \pm 1.508	90.352 \pm 2.025
NL	55.665 \pm 2.403	82.677 \pm 7.025
NL-OPT	54.430 \pm 1.602	80.987 \pm 5.703
HNL	50.003 \pm 1.359	76.477 \pm 4.576

Table 4: Consistency error comparison

Dataset	Method	Consistency error			Total
		5min vs 15min	5min vs 60min	15min vs 60min	
Load($\times 10^{-4}$)	LSTM	5.973	7.882	6.598	20.452
	MLP	4.52	6.04	3.758	14.318
	NL	3.319	3.838	3.718	10.874
	HNL	0.066	1.702	1.144	2.912
Wind power	NL	6.426	7.581	7.389	21.396
	LSTM	2.007	5.986	5.944	13.937
	MLP	2.084	3.932	2.499	8.515
	HNL	0.036	0.294	0.177	0.507

Table 5: Operation cost comparison on day-ahead scheduling ($\times 10^5$ ¥)

Method	5min	15min	60min
LSTM	123.902 \pm 10.309	41.63 \pm 2.852	10.38 \pm 0.797
LSTM-BU	123.902 \pm 10.309	42.635 \pm 3.461	10.566 \pm 0.865
LSTM-OPT	117.941 \pm 6.292	40.619 \pm 2.127	10.056 \pm 0.533
MLP	123.426 \pm 8.96	39.654 \pm 3.745	9.127 \pm 0.72
MLP-BU	123.426 \pm 8.96	42.464 \pm 3.011	10.496 \pm 0.757
MLP-OPT	115.559 \pm 6.105	39.813 \pm 2.041	9.838 \pm 0.517
NL	119.641 \pm 5.98	40.515 \pm 2.91	9.977 \pm 0.594
NL-BU	119.641 \pm 5.98	41.257 \pm 2.017	10.204 \pm 0.504
NL-OPT	116.377 \pm 4.509	40.166 \pm 1.517	9.934 \pm 0.371
Persistence	135.372 \pm 0.0	46.49 \pm 0.0	11.483 \pm 0.0
HNL	107.526 \pm 8.78	37.252 \pm 2.99	9.785 \pm 0.766

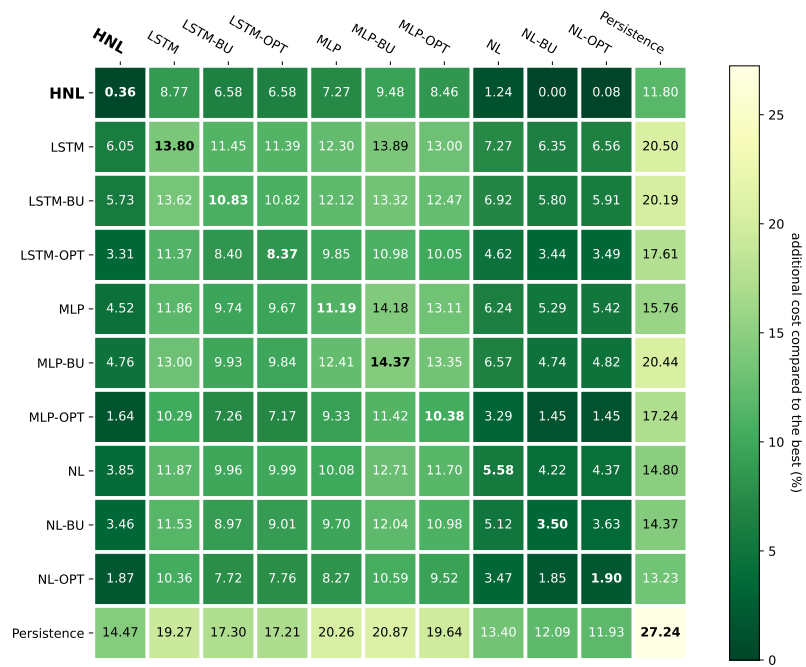


Figure 10: Comparison of operational costs for integrated scheduling (penetration rate: 20%)

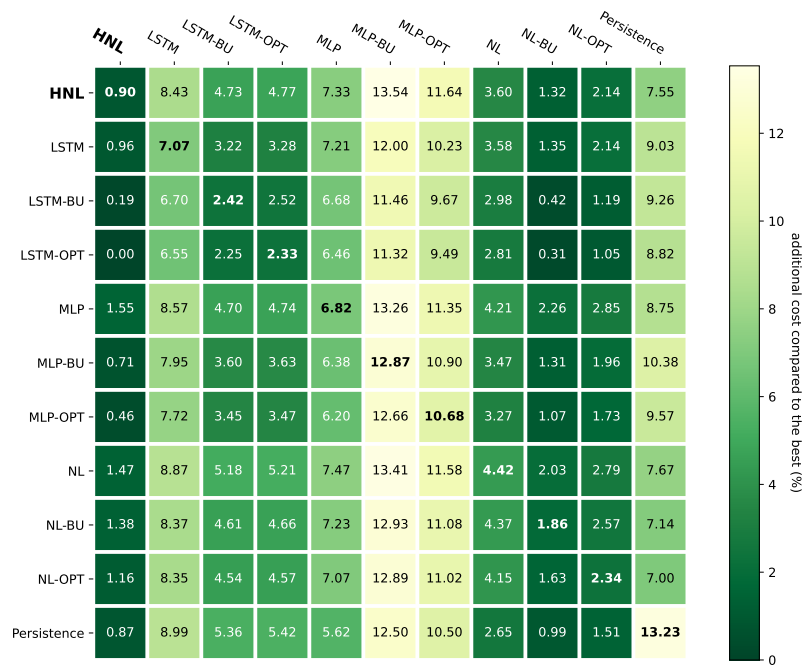


Figure 11: Comparison of operational costs for integrated scheduling (penetration rate: 80%)